university of groningen

University Medical Center Groningen

University of Groningen

Eyes on emergence

Nordhjem, Barbara; Petrozzelli, Constanza I. Kurman; Gravel, Nicolas; Renken, Remco J.; Cornelissen, Frans W.

Published in:
JOURNAL OF VISION

DOI:
10.1167/15.9.8

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2015

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Nordhjem, B., Petrozzelli, C. I. K., Gravel, N., Renken, R. J., & Cornelissen, F. W. (2015). Eyes on emergence: Fast detection yet slow recognition of emerging images. *JOURNAL OF VISION*, *15*(9), [8]. https://doi.org/10.1167/15.9.8

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 04-05-2023

Eyes on emergence: Fast detection yet slow recognition of emerging images

Barbara Nordhjem
Laboratory for Experimental Ophthalmology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Constanza I. Kurman Petrozzelli
Laboratory for Experimental Ophthalmology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Nicolás Gravel
Laboratory for Experimental Ophthalmology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Remco J. Renken
BCN Neuroimaging Center, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Frans W. Cornelissen
Laboratory for Experimental Ophthalmology, University Medical Center Groningen, University of Groningen, Groningen, The Netherlands

Visual object recognition occurs at the intersection of visual perception and visual cognition. It typically occurs very fast and it has therefore been difficult to disentangle its constituent processes. Recognition time can be extended when using images with emergent properties, suggesting they may help examining how visual recognition unfolds over time. Until now, their use has been constrained by limited availability. We used a set of stimuli with emergent properties—akin to the famous Gestalt image of a Dalmatian—in combination with eye tracking to examine the processes underlying object recognition. To test whether cognitive processes influenced eye movement behavior during recognition, an unprimed and three primed groups were included. Recognition times were relatively long (median ~ 5s for the unprimed group), confirming the object's emergent properties. Surprisingly, within the first 500 ms, the majority of fixations were already aimed at the object. Computational models of saliency could not explain these initial fixations. This suggests that observers relied on image statistics not captured by saliency models. For the primed groups, recognition times were reduced. However, threshold-free cluster enhancement-based analysis of the time courses indicated that viewing behavior did not differ between the groups, neither during the initial viewing nor around the moment of recognition. This implies that eye movements are mainly driven by perceptual processes and not affected by cognition. It further suggests that priming mainly boosts the observer's confidence in the decision reached. We conclude that emerging images can be a useful tool to dissociate the perceptual and cognitive contributions to visual object recognition.

Introduction

Object recognition is at the juncture of perception and cognition. Traditionally, there have been two approaches to the study of object recognition; either the emphasis has been placed on perceptual processes such as object detection and figure-ground segregation, or the focus has been on more cognitive aspects such as categorization and memory (Palmeri & Gauthier, 2004). Studying the processes underlying object recognition is challenging because visual recognition usually happens with seemingly little effort and is near instantaneous

Citation: Nordhjem, B., Kurman Petrozzelli, C. I., Gravel, N., Renken, R. J., & Cornelissen, F. W. (2015). Eyes on emergence: Fast detection yet slow recognition of emerging images. *Journal of Vision*, *15*(9):8, 1–16, doi:10.1167/15.9.8.

doi: 10.1167/15.9.8 Received September 30, 2014; published July 22, 2015 ISSN 1534-7362 © 2015 ARVO

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018

(Biederman, 1972; Potter, 1975; Schendan, Ganis, & Kutas, 1998; Thorpe, Fize, & Marlot, 1996).

The rapidity of visual recognition makes it relatively hard to examine how the progression from retinal signals to recognition of a meaningful object unfolds over time. However, the recognition process can be extended and postponed considerably by using images with emergent properties. The textbook example of such an image is the Dalmatian in a sun-spotted garden by photographer R. C. James (Figure 1). At first, the image just appears to consist of black spots, but eventually the dog will stand out from the background. The extended recognition times for such images allows for the use of eye tracking to study the gaze behavior before and after recognition and may provide insight into fundamental aspects of object recognition (Pelli et al., 2009). Images with emergent properties illustrate one of the main ideas of the Gestalt school, namely that perception is holistic. Indeed, the individual features of emergent images are practically unidentifiable when seen in isolation (Figure 1) indicating that recognition of global shapes precedes identification of individual parts (Wagemans et al., 2012). Visual emergence also demonstrate how the ability to recognize objects in a holistic manner rather than by grouping of individual parts is crucial for the flexibility of human object recognition (Kubilius, Wagemans, & Op de Beeck, 2011; Lee & Beeck, 2012).

The Dalmatian and a few similar stimuli were based on rare photographs, and until recently the amount of images with emergent properties has been rather limited because there was no systematic way to produce comparable stimuli (Ishikawa & Mogi, 2011). Recently, a new computerized method to synthesize stimuli with emergent properties was developed (Mitra, Chu, Lee, & Wolf, 2009). This technique derives stimuli—*emerging images* (EIs; Figure 2)—from 3D models in a systematic manner. The EIs were conceived specifically to provide as little information as possible for automated image recognition algorithms (Mitra et al., 2009). Yet, they can usually be recognized after a while by most human observers.

In the current study, the goal was to dissociate perceptual and cognitive contributions to visual object recognition by using computer-generated EIs. We did so by presenting viewers with EIs and focusing on recognition performance, viewing strategies, the influence of saliency, and the effect of priming. While observers attempted to recognize the images, we recorded their eye movements to study gaze behavior over space and time. During the viewing of natural images, eye movements are typically drawn first to the areas that stand out the most, their salient parts (Koch & Ullman, 1985). Hence, for the EIs, we also expected that early fixations would be more driven by salience compared to fixations made around the moment of recognition. Moreover, it is also not quite clear to what extent saliency-guided behavior may contribute to the recognition of EIs.

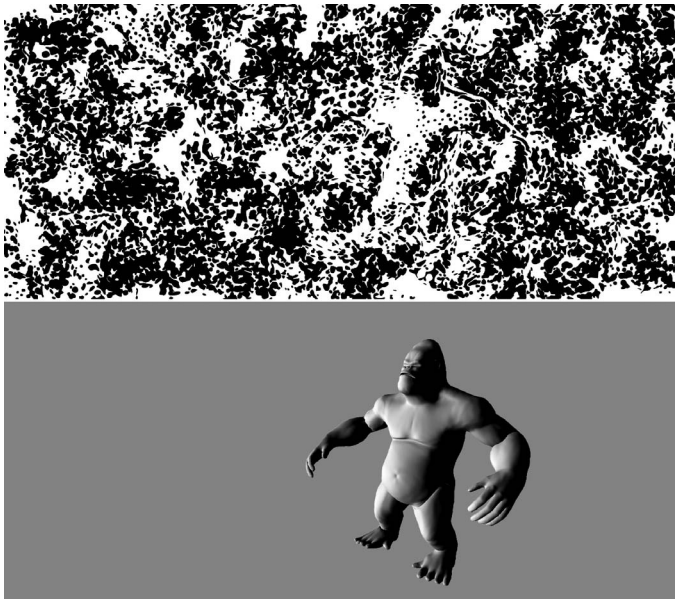

Figure 2. Example of an emerging image and the 3D model it was derived from.

We expected that during task performance, observers form hypotheses about the content of the image, which they will test by gazing in particular at potentially informative parts of the image (Geisler & Cormack, 2011). In other studies, eye tracking has

Figure 1. The Dalmatian by R. C. James (left), the same image with the dog highlighted from the background (center), and parts of the Dalmatian shown separately (right).

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018

revealed two spatio-temporal viewing strategies during the observation of visual scenes (Marsman, Renken, Haak, & Cornelissen, 2013; Pannasch & Velichkovsky, 2009; Unema, Pannasch, Joos, & Velichkovsky, 2005; Velichkovsky, Joos, Helmert, & Pannasch, 2005). Because of the extended recognition time, eye movements may reveal whether similar distinguishable strategies accompany the recognition of EIs.

Finally, we also included different types of priming to investigate the effect of cognitive processes on recognition time, accuracy, and eye movement behavior (Graf & Schacter, 1985; Schacter, 1992). Priming may also help to dissociate between different theoretical frameworks of visual representation (Biederman & Cooper, 1991; Marsolek, 1999). We expected that priming would result in faster recognition and higher accuracy (Biederman & Cooper, 1991; Fiser & Biederman, 2001; Malcolm & Henderson, 2009). We also specifically addressed the question whether different types of priming would result in distinctive eye movement behavior, which might indicate the perceptual and cognitive contributions to recognition.

Methods

Apparatus

All experiments were programmed in Matlab using the Psychtoolbox (Brainard, 1997) and the Eyelink Toolbox extensions (Cornelissen, Peters, & Palmer, 2002).

The stimuli were presented on a 22-in. CRT screen (LaCie Electron 22blue IV) with a resolution of 1920 × 1440 pixels and a refresh rate of 75 Hz. The screen had a luminance of black (0.1 cd/m^2), gray (55.5 cd/m^2), and white (104 cd/m^2). A remote eye tracker (EyeLink 1000) was used to track the eye movements of all participants. Calibration and validation of each individual participant was performed using built-in routines of the EyeLink software. Participants were seated in front of the screen, with their heads resting in a headstand and a viewing distance of 60 cm.

Stimuli

Fifteen EIs and eight similarly textured nonsense images were used as stimuli. Images were 1897 × 842 pixels, corresponding to an angular image size of 36.4° × 16.4°. The hidden objects in the EIs were animals and had an average size of 558 × 544 pixels, which corresponds to an average angular object size of 10.7° × 10.6°. The objects were all relatively large to ensure that subjects did not have to search for them due to their size. All hidden objects were shown from an iconic perspective and were placed at varying locations within the image. We ran a separate pilot study with 35 participants to select a set of stimuli that could be recognized by most of the observers (90%) and of which each image would take approximately the same amount of time to be recognized. Images that took observers on average less than 3 s or more than 10 s to recognize correctly were excluded. None of the participants who participated in the pilot study were included in the current experiment.

EI image generation

The EIs were generated from 3D models of animals (Figure 2). For a detailed description of the EI generation process and algorithm we refer to the conference proceeding by Mitra et al. (2009) and the project webpage (http://vecg.cs.ucl.ac.uk/Projects/SmartGeometry/emergence/emergence_image_siga_09.html). In short, the algorithm calculates an *importance map* based on the geometry, lighting, and view position. The importance map is constructed upon the silhouette and shading information of the object. The synthesis algorithm of the program turns the 3D model into *splats*, which texturize the image. These splats are scattered in such a way that they respect the features of the hidden object: shape, pose, and silhouette. Several parameters can be adjusted in the program. When generating the images, we focused on adjusting the density of the splats, splat size, and also making sure that the silhouette surrounding the hidden object was perturbed and not clearly distinguishable. The background clutter for each EI was copied and pasted by the algorithm from the splats comprising the object.

The EIs were derived from the same 3D models as used by Mitra et al. (2009) in their study. The precise parameter settings varied per image, with silhouette perturbation < 0.5, splat density ≈ 1.2, and perturbation displacement ≈ 0.005. A set of nonsense images was created using the *GNU Image Manipulation Program (GIMP)*. From the initial EIs, the areas with random splats were cropped, copied, and pasted on top of the hidden object to cover it. Following this procedure, the "paintbrush" tool was used to retouch any borders, making sure there was continuity in all the splats.

Participants

There were 67 participants who took part in the experiment, all of them with normal or corrected-to-normal vision. Ages ranged from 18–30 years. They were all naïve to the EIs and to the purpose of the experiment. All participants recruited for the study understood the instructions and were able to recognize an example EI.

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018

Priming and groups

To dissociate between different theoretical frameworks of visual representation (Biederman & Cooper, 1991; Marsolek, 1999), we utilized primes with the same shape, primes showing a different exemplar of the same object, and word primes. Participants were randomly assigned to one of four groups that were evaluated in this experiment. Priming was done with separate groups, because each EI can be shown only once per participant. All primes were presented at the center of the screen. The hidden objects in the EIs were placed at varying places, to avoid that the primes would simply cue location. The four groups were (a) Unprimed: Participants were not shown a prime, only a gray screen with a central fixation point prior to the EI (19 participants). Slightly more participants were assigned to the Unprimed group because we anticipated they would recognize less images based on the pilot study. (b) Same-shape: Participants were primed with a grayscale rendering of the 3D model used to create the EI (16 participants). The rendering had the same shape and size as the object hidden in the EI but did not give a location cue as all primes were all presented at the center of the screen. (c) Different-shape: Participants were primed with a grayscale photo of the same visual category as the object in the EI, but with a different shape and presented at the center of the screen (16 participants). (d) Word: Participants were primed with a written word naming the object in the EI (16 participants).

Procedure

The instructions within the overall experiment were to look at the EI and click with the leftmouse button if they recognized an object. Participants reported which object was seen by verbal response immediately after recognition was indicated. Subjects were instructed to indicate recognition when they saw an object they could name and categorize. Recognizing "something" or "an animal" was not considered specific enough. Naming an animal from a different class, such as a bird or a fish if the hidden animal was a mammal, was considered an incorrect response. In cases where an animal from the same class or with the same shape was recognized, we kept a print of each image and would let the subject trace the outline of the animal and describe where the different parts were perceived after the experiment. If they outlined the shape and were able to indicate where they perceived the different parts of the animal correctly, then the EI was considered as recognized. For the primed groups, the prime would already give away the correct answer. To circumvent the possibility that people would report recognition regardless of whether an object was actually recognized or not, we included the nonsense images. In the case of the nonsense images, a prime would still be shown, but it would be a prime randomly selected from the other primes in the same group.

Primes were presented for 1 s, and each EI was presented for 20 s. As a control condition, the corresponding rendering was presented for 10 s. Subjects had to respond to the sound of a bell that was played at a random time when the model rendering was shown. This task was included to measure reaction times and possible changes in eye movement behavior due to performance of the key press. The interstimulus interval was 1 s where instructions were shown on a gray background (instructions were "recognize" for the EI and "respond" for the model rendering of the hidden object viewed against a uniform background. Each trial lasted approximately 35 s (Figure 3).

Eye movement recording and preprocessing

Eye movements were recorded with an SR Research Ltd. Eyelink 1000 eye tracker with a sampling rate of 1000 Hz. A 9-point calibration was carried out followed by validation, also using a 9-point grid. Calibrations would be repeated until a spatial accuracy of $\pm$ 0.5° was reached. Drift correction was carried out prior to the presentation of each EI using a central fixation point. Fixations and saccades were parsed online using the algorithm provided by SR Research. The saccade velocity was set to a conservative threshold of 35°/s and acceleration to 9500°/ s^2. The data was processed off-line by excluding fixations made outside the image area and saccades starting or landing outside the image area. Fixations and saccades that were made between where the images appeared were excluded from the analysis. Trials where there were several jumps between fixations exceeding 10° of visual angle around the moment of key presses were also excluded.

Analysis of response time and recognition performance

The amount of correctly recognized images was compared between priming groups. Correct recognition was defined by naming the exact object or an object with a similar shape, which could be traced successfully on a print of the EI immediately after the experiment (see Procedure). Further, recognition times indicated by key presses were compared between groups. Not all variables were normally distributed in all groups. Therefore, we report the median (*Mdn*) and interquartile range (IQR). The nonparametric Kruskal-Wallis test was used to test the main difference between

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018

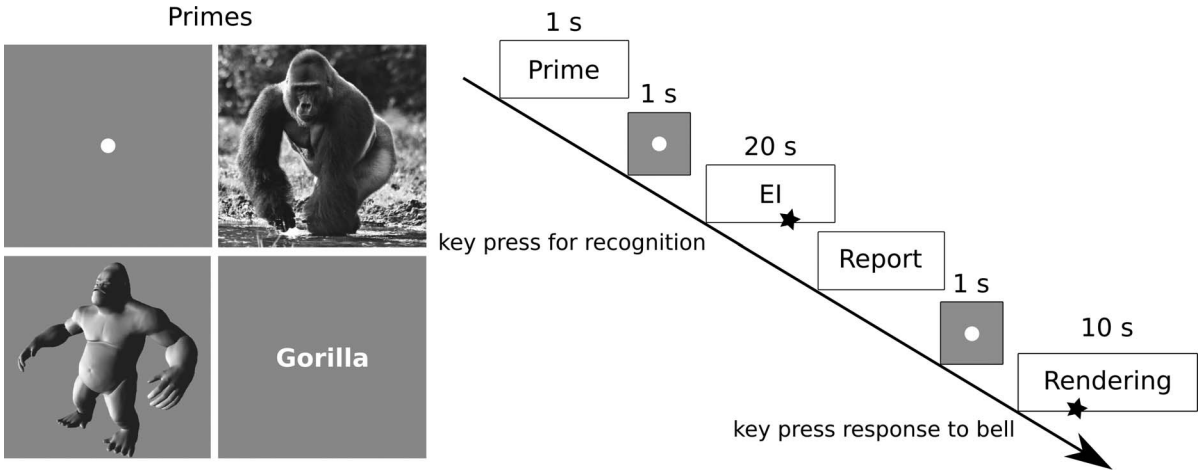

Figure 3. Each participant took part in one experimental run. Before the experiment, participants were shown the image of the Dalmatian and the task was explained. An experimental run consisted of 23 trials. The presentation order in a trial was Prime, Central fixation point, EI, Verbal report, Rendering.

groups, and pairwise Mann-Whitney tests, corrected for multiple comparisons, were carried out to compare groups. Statistical tests were computed in SPSS.

Fixation maps

The iMap3 toolbox (Caldara & Miellet, 2011) was used to create fixation maps. Fixation maps are based on coordinates of fixation locations (x, y) across time, and weighted by fixation durations. The resulting fixations distributions were smoothed with Gaussian kernels with a standard deviation of 10 pixels. The fixation maps of all observers were summed together separately for each EI. The maps were used to visualize where observers were fixating for the first 1000 ms of image viewing and for 1000 ms before the moment of recognition.

Analysis of eye movements over time

The time courses of fixation durations and saccade amplitudes were plotted from the onset of the EIs. To investigate how viewing behavior changed around the moment of recognition, data was also centered on the moment of recognition. In order to examine the role of perceived edges in recognition, we calculated Euclidean distances of fixations to the nearest edge of the object for each image. Edges were defined by extracting the outlines of the model renderings from which the EIs were derived. Thus, a region of interest (ROI) was defined individually for each EI. Distances were initially found in pixels and then converted to degrees of visual angle. Distances were defined relative to the edge of each ROI, with negative values being outside and positive values being inside the object. It is possible that some of our observations are not due to the EIs but reflect certain biases; participants may, for instance, be more likely to look at the middle of the screen (Bindemann, 2010; Tatler, 2007). To test the null hypothesis, that there was no relation between fixations and edges around the moment of recognition, we randomly paired fixations and objects over 10 iterations. Thus, any patterns due to just viewing images over a period of time but not related to recognition of a particular object should be visible when plotting the random pairings. To investigate dynamics of viewing behavior around the moment of recognition, we plotted fixation duration and saccade amplitude. For all parameters, the median for each time bin was plotted with the interquartile range as well as the 90 % range. We opted for the median and not the mean because the data was highly skewed.

We compared the eye movement time courses of the four priming groups from trial onset and around the moment of recognition, and trials where the object was recognized, with trials where recognition did not occur in terms of eye movement behavior using the same approach. Comparisons of time courses were carried out by implementing a modified version of threshold-free cluster enhancement (TFCE; Smith & Nichols, 2009). TFCE has the advantage of both optimizing detection of smaller signal changes that are consistent in time as well as sharp peaks. TFCE scores represent the supporting data under the curve, taking both height as well as temporal continuity into account. Hence, TFCE integrates duration and effect size of a response into a single statistic for each time point. TFCE was initially implemented for fMRI research data but has also been adapted for comparison of fixation maps (iMap3; Caldara & Miellet, 2011) and EEG data

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018

(Mensen & Khatami, 2013; Pernet, Chauveau, Gaspar, & Rousselet, 2011). Distance to edge, fixation duration and saccade amplitude was compared by calculating TFCE difference values between groups to investigate if priming had an effect on viewing behavior. The TFCE difference values were compared for the median, the 5th, and 95th percentile. Significance values were obtained using permutation statistics (1000 permutations) with a correction for multiple comparisons across groups ($p < 0.05$). Further, three uncorrected comparisons ($p < 0.05$) were made contrasting Unprimed with each of the three priming groups (see S2 for the TFCE parameters).

Predicting fixations using models of saliency

Saliency maps were computed for all EIs to determine whether fixations were guided by salience. We used two computational models of salience; the classic saliency model (Itti, Koch, & Niebur, 1998), and the Graph-Based Visual Salience (GBVS) model (Harel, Koch, & Perona, 2006). We used the "Graph-Based Visual Saliency" Matlab toolbox by J. Harel, which included both saliency models tested. For the sake of comparison, we also calculated salience for the model renderings. We assessed the predictions of both saliency models, comparing the probability of hits and false alarms using the Receiver Operating Characteristic (ROC) metric and reporting the area under the curve (AUC). The greater the AUC, the better the model discriminates between correct and false model predictions. The ROC curve can be summarized by its AUC where 0.5 corresponds to chance (a linear line), and 1.0 corresponds to a perfect discrimination. To test how well saliency models predicted fixations against the null hypothesis, we used random pairings of images and eye movements over 10 iterations per image. We used random pairings of EIs and fixations instead of just generating random fixation coordinates to ensure that we take general tendencies such as center bias into account. We calculated the ability of salience models to predict fixations made on EIs, compared with random pairings of fixations and EIs, using paired t tests.

Results

We recorded recognition times and eye movements during recognition of EIs to study the perceptual and cognitive processes involved in visual object recognition. Surprisingly, most participants detected the emergent objects within 500 ms with their eye movements while recognition was indicated later in time; this fast detection was also found for EIs that were not recognized at all. Eye movements were not guided by salience: Both the classic and the more recent saliency model could not predict fixations or the location of the hidden objects. Priming affected recognition time, but not gaze behavior. Below, we will describe these findings in more detail.

Comparison of recognition performance over priming groups

We expected that all primed groups would show faster recognition times and higher accuracy than the unprimed group. Based on previous studies (Biederman & Cooper, 1991; Fiser & Biederman, 2001), we expected that primes with the same shape would be most effective in reducing recognition time and improving accuracy, and that primes showing the same object category would be more effective than word primes (Malcolm & Henderson, 2009).

In all four priming groups, the majority of the participants successfully recognized most images (Figure 4). In the Unprimed group, $Mdn = 80$ % (IQR 73.3%–93.3%) of the EIs were recognized. The highest percentage of recognition was obtained in the Same-shape primed group $Mdn = 100$ % (IQR 93.3%–100%), while $Mdn = 93.3\%$ (IQR 83.3%–100%) were recognized in the Different-shape primed group and $Mdn = 90\%$ (IQR 86.7%–100%) in the Word-primed group. The Kruskal-Wallis test was used to compute the main effect and Mann-Whitney pairwise comparison tests, adjusted for multiple comparisons, was carried out between the priming groups. There was a significant main effect of priming on the number of images recognized, $H(3) = 17.43$, $p < 0.05$. The only significant pairwise comparison was between the Unprimed and the Same-shape group ($U = -23.33$, $r = -0.69$, $p < 0.001$). None of the other groups differed significantly from each other.

Furthermore, we analyzed recognition times based on the moment of key press (Figure 5). The longest recognition times (RTs) occurred for the Unprimed group ($Mdn = 4800$ ms, $IQR = 2600 - 8400$ ms), while the shortest recognition times were found for the Same-shape primed group ($Mdn = 1600$ ms, IQR 1100–2800 ms). Similar RTs were found for the Different-shape primed group ($Mdn = 2500$ ms, IQR 1400–4900 ms) and for the Word-primed group ($Mdn = 2400$ ms, IQR 1400–4600 ms). The response time to the bell sound during viewing of the rendering following each EI across groups was also calculated ($Mdn = 720$ ms, IQR 503.8–824.5 ms).

There was a significant main effect of priming on RT, $H(3) = 149.37$, $p < 0.05$. Mann-Whitney tests were carried out to compare the groups with p values adjusted for multiple comparisons. RTs in the Unprimed group

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018

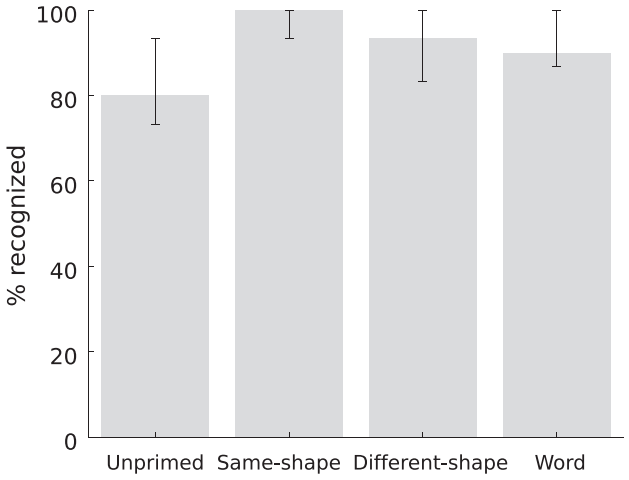

Figure 4. Median recognition accuracy for emerging images for the various types of priming. Error bars indicate the interquartile range.

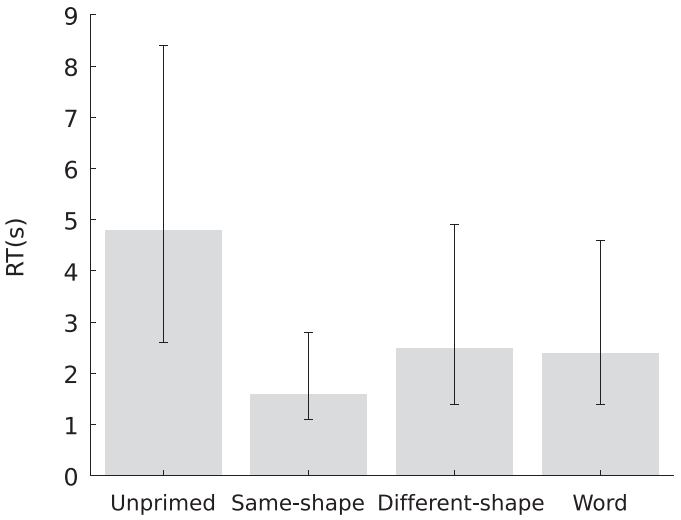

Figure 5. Median recognition time (RT) for emerging images for the various priming groups. Error bars indicate the interquartile range.

was significantly different ($p < 0.001$) from the Same-shape ($U = 280.03$, $r = 0.57$), the Different-shape ($U = 154.03$, $r = 0.31$) as well as the Word-primed group ($U = 0.33$, $r = 0.33$). RTs in the Same-shape group differed significantly ($p < 0.001$) from RTs in both the Different-shape ($U = -126.0$, $r = -0.26$) and the Word ($U = -118.16$, $r = -0.24$) group. The Different-shape and Word-primed groups did not differ significantly from each other ($U = 7.84$, $r = 0.02$). Hence, the results show that the priming did have an effect, and the most effective primes were the Same-shape images.

Fixation maps

To spatially examine viewing behavior, we computed fixation maps while aligning the trials based either on the start of the trial or on the moment of recognition. For participants who eventually recognized the object, we computed fixation maps for the first and second 500 ms bin, as well as for the 1000 ms preceding the moment of recognition. Figure 6 shows fixation maps for the Gorilla EI. The fixation map in Figure 6A indicates that most participants managed to locate the object already within the first 500 ms of viewing the image. Note, however, that observers were primarily looking at the chest rather than at the head. In the second 500 ms bin, most fixations were on the head (Figure 6B). Around the moment of recognition, the head was primarily fixated (Figure 6C).

Does saliency predict the fixation locations?

Given the fast detection of the object location within the EIs, it is reasonable to wonder whether visual saliency might predict this behavior. For this reason, we investigated how well a classic (Itti et al., 1998) and more recent (Harel et al., 2007) salience model predict the fixations (in the following the models are referred to as Itti and GBVS saliency, respectively). Salience models predict the conspicuous features in an image that will attract gaze based on image characteristics such as luminance, contrast, orientation, and color. Generally, low predictive power of the saliency maps was expected, given that several computer vision algorithms have failed to characterize the objects hidden in the EIs (Mitra et al., 2009). In order to evaluate the agreement between saliency maps and a set of fixations made on the image, we computed an AUC score for each image where chance level is 0.5, and perfect prediction is 1.0. We compared the AUC scores for EIs and for the renderings, and for random pairings of images and sets of fixations.

Since it is possible that initial fixations are guided more by salience than later ones, we carried out two separate analyses. We carried out a salience analysis for fixations made within the first 1000 ms of image presentation, as well as one for fixations made in a 1000 ms window centered on the moment of recognition of the EIs. As a control, we performed the same type of analysis using the fixations made within the first 1000 ms of presenting the rendering, and for a 1000 ms window centered on the moment of the key presses made during the presentation of the model renderings.

Results are shown in Table 1, and saliency maps are shown for an EI in Figure 7. Generally, the AUC scores were higher for the GBVS than for the Itti saliency. Not surprisingly, both saliency models performed well for the fixations on the renderings, and performance

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018

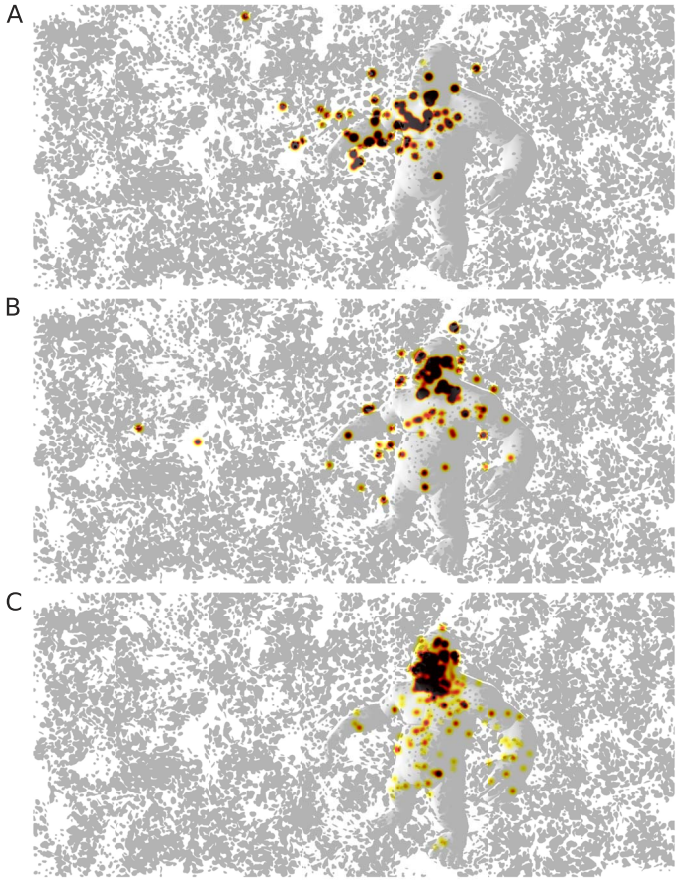

Figure 6. Fixation maps with trials aligned on the start of the trial. (A) The first 500 ms. (B) 500–1000 ms. Only data for participants who eventually recognized the gorilla EI is included in this map. (C) Fixation map with trials aligned with respect to the moment of recognition. The map shows the fixations that occurred during the 1000 ms prior to the moment of recognition. Note that for illustration purposes, the model rendering is superimposed on the EI (the actual EI is shown in Figure 2).

		Mean AUC	*SD*	*t*	*p*
Saliency for EIs					
From onset	Itti	.49	.09	−.73	.48
	$Itti_{rand}$	.50	.07		
	GBVS	.82	.05	.20	.84
	$GBVS_{rand}$	.82	.2		
Centered on recognition	Itti	.53	.08	−.30	.77
	$Itti_{rand}$	.53	.05		
	GBVS	.76	.08	.22	.83
	$GBVS_{rand}$	.77	.02		
Saliency for renderings					
From onset	Itti	.87	.02	11.37	<.001
	$Itti_{rand}$	.64	.08		
	GBVS	.89	.02	10.21	<.001
	$GBVS_{rand}$	.66	.08		
Centered on recognition	Itti	.90	.03	12.16	<.001
	$Itti_{rand}$	.66	.07		
	GBVS	.92	.03	11.42	<.001
	$GBVS_{rand}$	.67	.09		

Table 1. Ability of saliency maps to predict eye movements for EIs, renderings, and random pairings of images and fixation locations over 10 iterations per image. Paired *t(14)* tests showing the difference in how well salience maps predict eye movements by comparing AUC scores for each image with the null hypothesis, namely random parings of images and eye movements (Denoted $Itti_{rand}$ and $GBVS_{rand}$; 10 iterations per image). The predictive power of saliency maps was both calculated for EIs and for renderings of the objects they were derived from.

decreased substantially for random pairings. In contrast, for neither the early nor the later fixations made during the presentation of the EIs, the saliency models performed better for the actual than for the random pairings of fixations and images. This result shows that salience is not a good predictor of the fixations made on EIs, suggesting that the low-level visual features captured by the saliency models do not guide the eye movements to the objects.

Finally, it may be possible that initial fixations are more guided by salience for some types of priming compared to others. To compare whether saliency models differed in predictive power across priming types, we carried out an ANOVA. There were neither differences between the AUC scores for the priming groups during initial viewing for Itti saliency, $F(64, 3) = 0.05$, $p = 0.98$, nor for GBVS saliency, $F(64, 3) = 0.4$, $p = 0.76$. Hence, there is no evidence that priming affects to what extent initial fixations were guided by salience.

Temporal analysis of eye movement behavior from trial onset

We plotted the median distance to the nearest edge, fixation duration, and saccade amplitude within a time window starting at trial onset and ending 2000 ms later (Figure 8). The distance-to-nearest-edge plot also shows the null hypothesis based on random pairings of EIs and eye movements over several iterations (Figure 8A). This way, there are the same temporal patterns in the eye movements in the null hypothesis. If there is a spatial bias, such as fixating more on the center of the screen, this will also be preserved, while the spatial relation between fixations and the hidden animals in the EIs is disrupted. For each plotted parameter, the darker shaded area shows the interquartile range whereas the lighter shaded area shows the 90% range. The distance-to-the-nearest-edge plot shows that the gaze of the observer approached the objects' edges after

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018

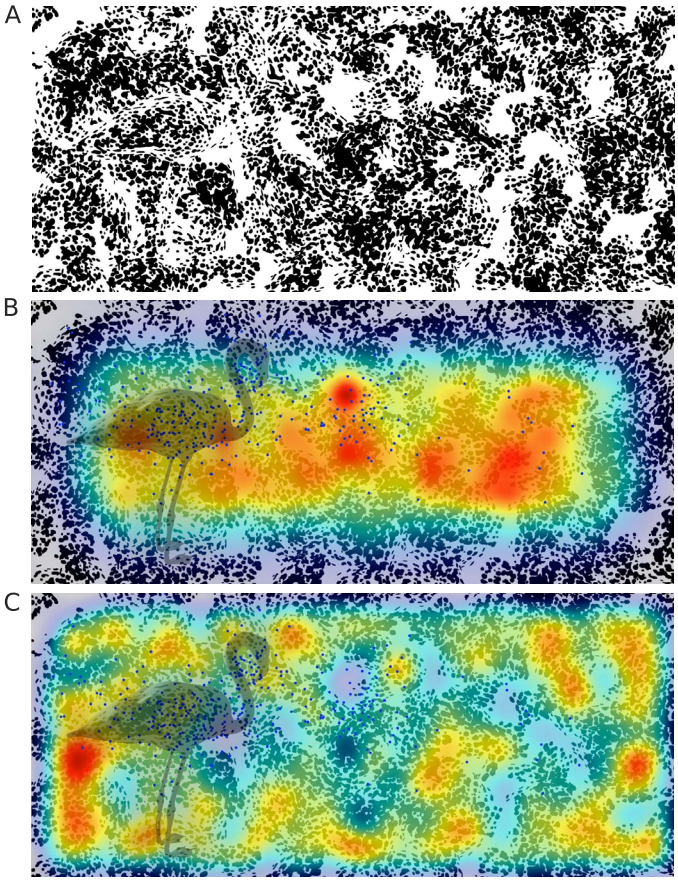

Figure 7. Saliency maps computed for the flamingo EI (A) with Itti saliency (B) and the GBVS (C) algorithm. The fixation locations for the first 2 s of viewing across groups are shown with blue dots; the hidden object (flamingo) is shown in a darker shade here for illustration purposes.

the first 500 ms, at which point the median and interquartile ranges reach a plateau and become stable (Figure 8A). Median fixation duration increased after approximately 500 ms and hereafter became relatively stable (Figure 8B), while median saccade amplitude decreased within the initial 500 ms (Figure 8C). Saccade amplitude plotted as a function of fixation duration showed the largest saccade amplitudes and can be observed for fixations with duration of 80–120 ms (Figure 8D).

Temporal analysis of eye movement behavior centered on the moment of recognition

Figure 9 shows the median viewing behavior over all groups in a 4000 ms temporal window centered on the moment of recognition. The darker shaded area shows the interquartile range, whereas the 90% range is shown by lighter shading. Overall, around the moment of recognition, the distance to the nearest edge of the fixation positions showed little change in the median and interquartile range. However, there was more variation in the 90% range: 2000–1000 ms prior to recognition, part of the fixations landed at relatively large distances to the edge. Around 1000–500 ms prior to recognition, one can observe a marked decrease in variability in this behavior. Median fixation duration also increases slightly prior to recognition and remains higher from that moment onwards. This increase in fixation duration is accompanied by an increase in variability as well. Saccade amplitude (Figure 9C) does not show any marked changes around the moment of recognition. Figure 9D plots saccade amplitude as a function of fixation duration. The data follows a similar trend to the data shown in Figure 8D. Saccade amplitude shows somewhat of a peak for fixations that last around 100–120 ms and is lower for fixations that last either shorter or longer than this.

Comparison of viewing behavior between recognized and unrecognized trials

We compared trials in which successful recognition took place with trials in which participants did not recognize an object using the TFCE analysis and permutation statistics. The time courses compared spanned over 2 s from trial onset. We found no significant differences between successfully recognized and unrecognized trials for distance to edge, or saccade amplitude using a threshold of $p < 0.05$ uncorrected for either the 5^{th}, 50^{th}, and 95^{th} percentile per bin. However, the analysis revealed a significant difference in terms of fixation duration between recognized and unrecognized trials: After the initial 500 ms of viewing, fixation durations were longer for trials where an object eventually was recognized compared to trials where recognition did not occur. To illustrate this contrast, we have plotted the median fixation duration and the interquartile range (Figure 10).

Comparison of viewing behavior in different priming groups

Having found marked differences in reaction time in relation to priming, we wondered whether the priming would be apparent in different viewing behavior.

In order to statistically compare differences in viewing behavior over time between groups, a TFCE analysis was performed per time course and compared between groups using permutation statistics. The comparisons revealed no significant differences between priming groups for either distance to edge, fixation duration, or saccade amplitude using a threshold of $p <$ 0.05 uncorrected for either the 5^{th}, 50^{th}, and 95^{th}

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018

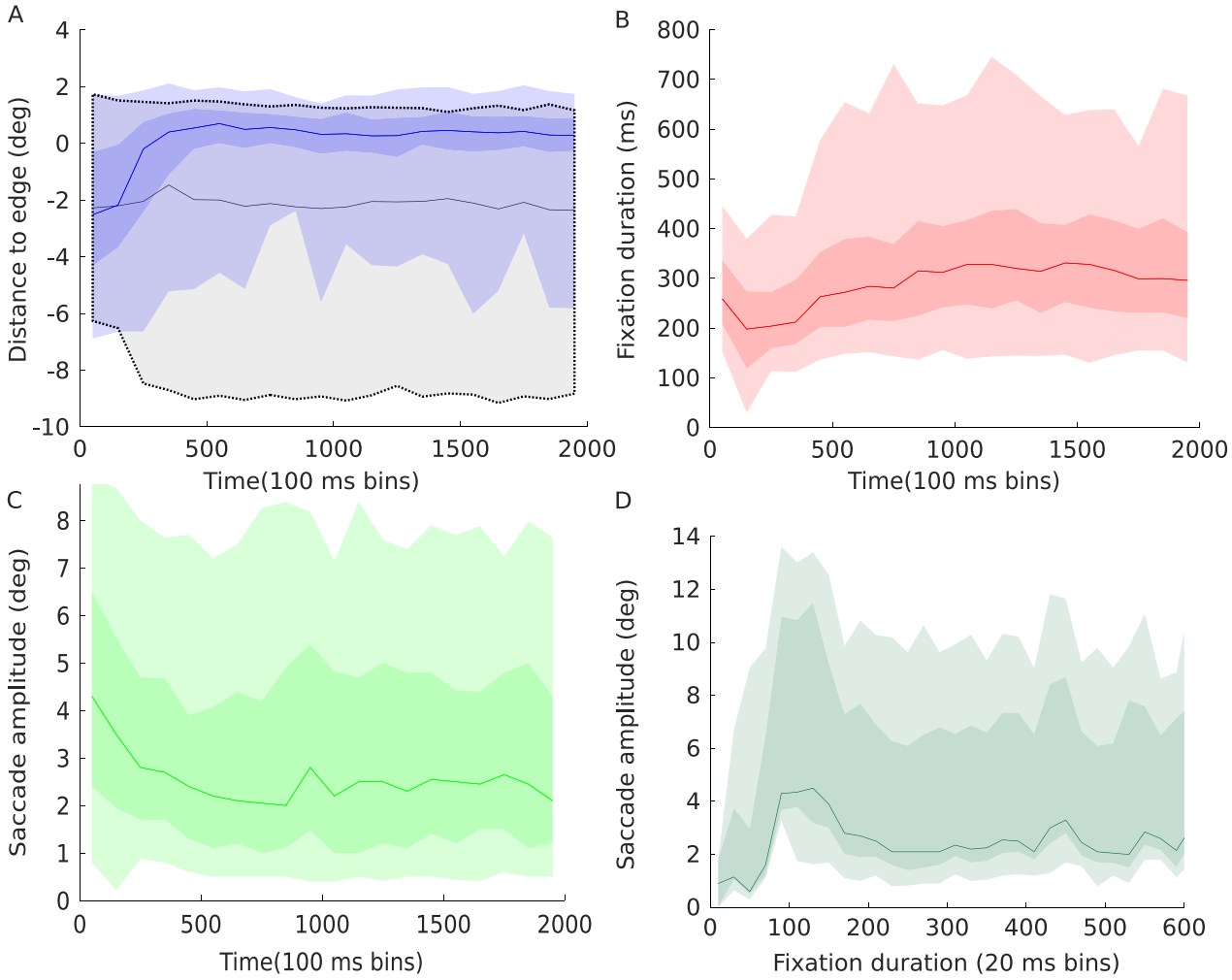

Figure 8. Viewing behavior during the initial 2000 ms of observing Emergent Images. (A) Distance to the nearest edge of the hidden object. The null hypothesis is plotted in gray. (B) Fixation duration. (C) Saccade amplitude. (D) Saccade amplitude plotted as a function of fixation duration.

percentile per bin. This was found for both time courses from trial onset and centered on the moment of recognition.

Discussion

We investigated the recognition of emergent images (EIs) by measuring recognition times and concurrent eye movements. Our main results are

- A new set of images with emergent properties was identified.
- Observers—who recognized the objects only after several seconds—were already looking closely at the object's position within the first 500 ms, indicating rapid detection of the hidden object's location.
- Salience did not predict fixations on the EIs, neither during initial viewing nor around the moment of recognition.
- Just prior to the moment of recognition, changes in viewing behavior were most apparent from the increased consistency with which observers gazed at the object. This behavior was accompanied by a concurrent increase in fixation duration. Saccade amplitude did not change notably during this time.
- Manipulating the available cognitive information by priming had an effect on recognition time but not on eye movement behavior around the moment of recognition. The Unprimed group and the three different priming groups (Same-shape, Different-shape, Word) did not show differences with respect to viewing behavior (median distance of fixations to the edges of the object, fixation duration, or saccade amplitude).

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018

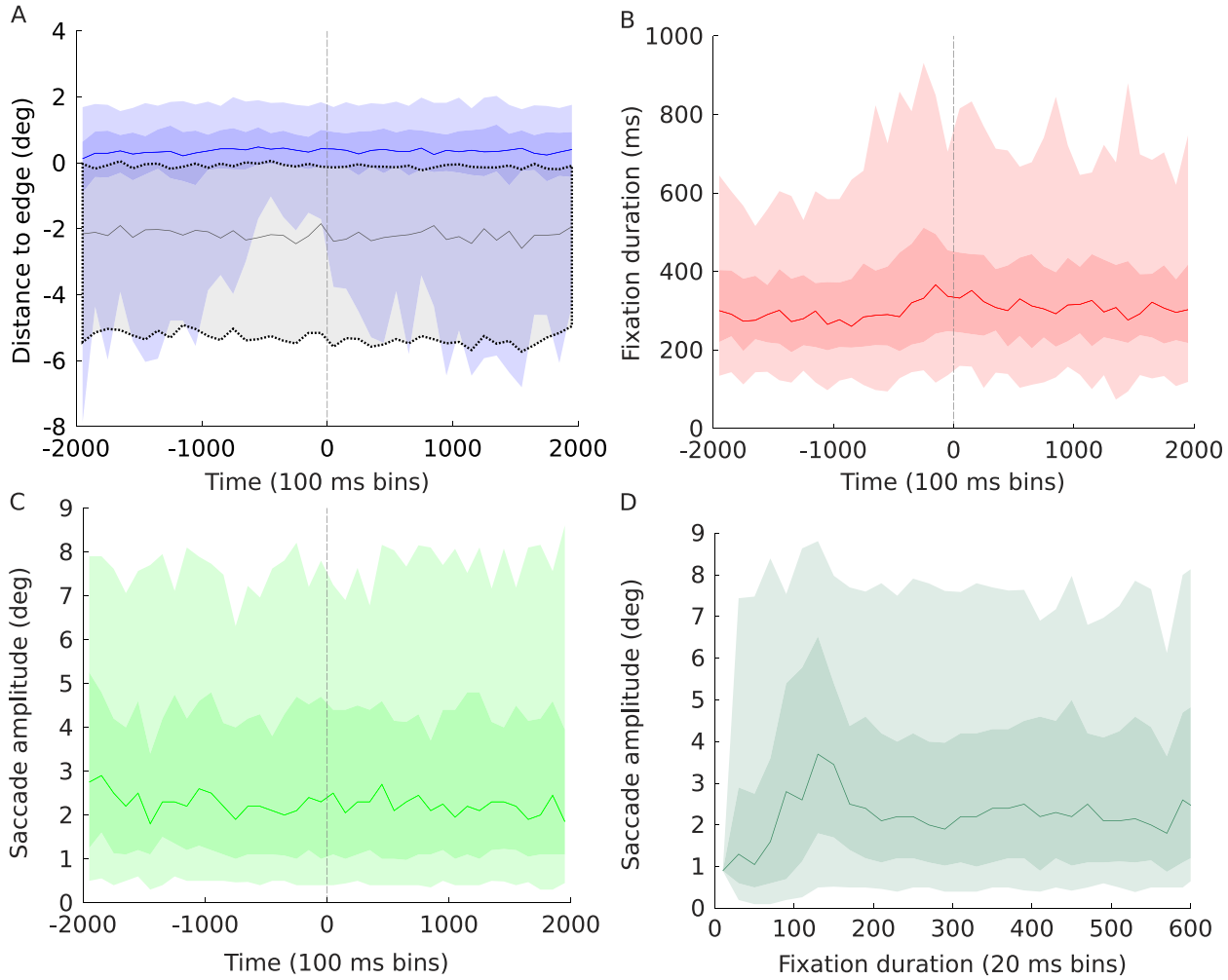

Figure 9. Viewing behavior centered on the moment of recognizing the content of the Emergent Images. (A) Distance to the nearest edge. The null hypothesis is plotted in gray. (B) Fixation duration. (C) Saccade amplitude. (D) Saccade amplitude plotted as a function of fixation duration.

Below, we will discuss these results and their implications in more detail.

A new set of emerging images has been identified

While the phenomenon of emergence has been used in the study of object recognition before, its use has been limited by the availability of only a few unique images that by now have been used for decades (the famous Dalmatian image was first published in *LIFE Magazine* in 1965). We generated a new set of stimuli using a computer algorithm developed by Mitra et al. (2009), which were subsequently evaluated for recognition time and performance. Note that not every image generated by the algorithm has automatic emergent properties for human observers. These images require verification and selection through measuring performance and recognition time. Based on our

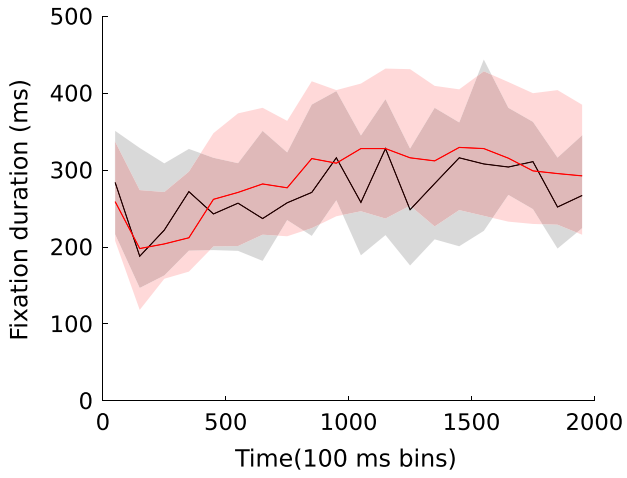

Figure 10. Fixation duration for recognized (red) and unrecognized (gray) trials plotted with the interquartile range.

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018

testing, we have now identified 15 new images that can be recognized successfully by nearly all observers yet still take several seconds to do so, thus indicating their actual emergent character. Having a much larger set of emergent stimuli available may contribute to future studies to understand the process of human visual object recognition. This type of stimuli could also be suitable for use in neuroimaging studies. Given the low temporal resolution of functional magnetic resonance imaging, stimuli that take a long time to recognize will be useful for examining the processes proceeding and underlying visual recognition. The identified set of images may also prove useful for evaluating future saliency and computer vision models, in particular those striving to closely mimic human vision.

Priming improves and speeds up recognition of EIs

Participants were assigned to either an unprimed group or three different priming groups to investigate the effect of cognitive processes on recognition performance. Confirming earlier priming work, compared to the unprimed observers, all primed observers recognized more EIs and required less time for recognition. Between the three priming groups, there was no difference with respect to the number of images recognized. Apparently, knowing which object to look for was sufficient to enable more observers to identify it and to do so more rapidly. There was no specific advantage of matching the same shape or having seen a visually similar image, since all priming groups performed equally well.

Priming also resulted in markedly shorter recognition times overall, but there were differences between the various primes. The observers primed by the Same-shape prime required less time to recognize the EIs than the observers in the Different-shape or the Word-primed group. The advantage that the Same-shape primes provided cannot be explained by location as all primes were presented at the center of the screen. This result corresponds with previous findings showing that Same-shape images are more effective than Different-shape and Word primes (Biederman & Cooper, 1991; Fiser & Biederman, 2001; Malcolm & Henderson, 2009). The type of priming has the same advantage on regular image as well as EIs, suggesting similar underlying cognitive recognition processes.

Fast detection, but slow recognition of emerging images

A possible explanation for the long recognition times of EIs would be that the EIs primarily extend the time required to find the object within the image while the recognition process itself is as fast as usual. First, to ensure that search was not the main task, we made the emerging objects relatively large. Hence, EIs were difficult to recognize due to their lack of conspicuous features rather than due to their small size. In addition, we found that within 500 ms, the subjects who eventually recognized the object were already gazing at it. This response indicates that the image region containing the object region was detected very rapidly upon presentation of the image and well before observers indicated recognition of the hidden object had occurred. Object search time was thus only a minor component of the recognition time.

Saliency models do not predict the fixations on EIs

A possible explanation for the fast detection of the object region would be that it stands out because it is more salient. For this reason, we analyzed how well a classic and a more recent saliency model predicted fixations made on EIs during initial viewing and around the moment of recognition (Harel et al., 2006; Itti et al., 1998). We found that neither of the models would clearly mark the image regions with the object as being more salient. Since human observers did fixate on and near the objects, the saliency models were also poor in predicting human fixation performance. This finding is in line with Mitra et al.'s (2009) demonstration that human observers exhibited superior recognition performance compared to three biologically-inspired vision algorithms (Epshtein & Ullman, 2005; Nister & Stewenius, 2006; Serre, Wolf, Bileschi, Riesenhuber, & Poggio, 1999). To our knowledge, at present there is no algorithm that can reliably detect the objects in the EIs.

Unknown image statistics makes the objects stand out for the human visual system

Saliency models—that emulate the early feature processing stages of human vision—fail to detect EIs and predict eye movement. However, participants fixated the region containing the object already within 500 ms. Such fast localization of the object region suggests that human vision extracts a statistic from the *splats* of the EI that makes the object stand out and attract gaze. In support of the idea that image statistics guide eye movements, we found that rapid eye movement directed at the objects, both by observers who did and did not eventually recognize the EI. This idea that specific image statistics are crucial for recognition has been shown in a previous study using the Dalmatian image (Tonder & Ejima, 2000). In that study, most participants could locate the bulging body of the dog, even though many were unable to correctly

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018

identify the object or its parts. However, when the experimenters changed the local texture orientation in the bulging body, most participants failed to detect the Dalmatian. Note that the saliency models we tested do compute orientation contrast, but apparently this fails to capture the relevant image property.

Viewing behavior: From scanning to inspection

Within the first second of viewing, we observed a transition from shorter to longer fixations and from larger to smaller saccade amplitudes. It is likely that the initial viewing phase was related to scanning the EIs and a brief search for the emerging object, whereas more close inspection followed, and eventually recognition. This finding came somewhat as a surprise to us; we had expected that observers viewing the EIs would require a longer period of scanning before inspection would take place. Within the initial period of viewing, saccade amplitude was between 3°–4°. Saccades made on EIs were shorter than those mostly observed in scene viewing, which are typically $> 5°$ (Over, Hooge, Vlaskamp, & Erkelens, 2007; Unema et al., 2005). This distinction suggests that observers exploited neural filters in parafoveal, rather than in peripheral vision to identify EI features. The lack of salient regions may also have dampened saccadic amplitude. The eye movement behavior we observed is similar to previous studies. Free viewing of scenes can be characterized by an initial period of spatial orientation—the ambient or scanning mode of attention—which after approximately two seconds is followed by more detailed inspection—the focal or inspection mode of attention (Marsman et al., 2013; Pannasch & Velichkovsky, 2009; Unema et al., 2005; Velichkovsky et al., 2005). Scanning is characterized by relatively large saccades and relatively short fixations, whereas during inspection saccades are smaller and fixations last longer, implying scrutiny of elements within the scene. Hence, eye movements on EIs largely resemble eye movements on regular images.

More consistent viewing behavior and longer fixation durations around the moment of recognition

We observed that fixation locations leading up to the moment of recognition were rather consistent both before and after recognition. There were no evident changes before or after the moment of recognition in fixation locations when plotting the median and interquartile range. The plot showing the median distance of fixations to object edges was literally a flat line. However, the tail of the distribution indicates that there were eye movements which targeted regions distributed over the whole image in time intervals further from the moment of recognition. This behavior changed around about 1000 ms before recognition; instead of targeting the background, fixations more often were targeted close to or inside the object. Hence, the 90% interval diminished and showed that more fixations were made closer to the objects. Over the same period of time, fixations became longer while saccade amplitude remained the same.

A possible interpretation of these results is that up to one second before recognition, observers had already identified a region that they considered most likely to contain the object. However, they were also looking at the background to consider other candidate areas. Around the moment of recognition, a change happened. Observers felt certain enough to indicate recognition, and were just focusing on information from the object. As outlined in the previous section, there is a transition from an ambient to a focal mode during initial viewing. At the moment of recognition, focal viewing behavior became even more pronounced; there was a moment of "hyper-focal" viewing with prolonged fixations on the object.

Detection and recognition of EIs is probably a complex process that relies on detection of structure, active hypothesis testing, and previous exposure (Lee, 2003). Active hypothesis testing—ultimately leading to recognition—is supported by the eye movement behavior we have measured. Participants were primarily looking at the object but continued probing the background before recognition, whereas at the moment of recognition, they fixated almost exclusively on the object.

Unrecognized objects were nevertheless detected rapidly

We compared viewing behavior for trials in which the EIs were successfully recognized with trials where participants did not indicate recognition. Interestingly, we did not find a difference in the distance to edge of fixations or saccade amplitude. Hence, this behavior indicates that for most observers, attention was guided towards the hidden objects regardless of whether an object eventually was recognized or not. This outcome again supports that there is something in the structure in the EIs that give away the objects: The right area was being detected, but participants may have lacked the confidence to decide exactly what they were looking at.

There was a difference between successfully recognized and unrecognized EIs during initial viewing. After the first 500 ms, fixation durations were shorter for unrecognized images compared to trials in which recognition did occur. We speculate that the increase in fixation duration is related to the observer's certainty

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018

that the right object has been detected; this would be consistent with our finding that fixation durations increased just around the moment of recognition. Therefore, the shorter fixation durations may reflect more uncertainty in the unrecognized trials.

Priming boosts confidence in decision making yet does not alter eye-movement behavior

It has long been known that the task influences how observers examine images (Buswell, 1935; Yarbus, 1967). Therefore, different viewing strategies for primed and unprimed groups could also be expected. However, we found that, unlike different viewing tasks, priming does not impact viewing behavior. A threshold-free cluster enhancement (TFCE) analysis of the scan paths showed that, for both initial viewing and eye movements around the moment of recognition, priming did not affect eye movement behavior. There were no differences between the groups in terms of fixation duration, saccade amplitude, and fixation distance to the nearest object edge. In the unprimed group, one could have predicted a larger degree of mislocalization of object boundaries, but the unprimed observers were not further from the object edges than the primed observers. These results support the point made in the previous section, namely that low-level features mainly guide eye movements. Our finding of invariant eye movement behavior under different priming regimes suggests that prior information does not impact the way that the EIs are viewed.

Priming resulted in faster recognition times. This could either be explained by faster localization, a more efficient testing of perceptual hypothesis, or greater confidence that the right object was recognized in the EIs. The first two explanations predict differences in eye movement behavior, whereas the latter does not. We did not find differences between groups regarding their eye movement behavior. Hence, we conclude that priming primarily affected observers' confidence, resulting in faster decision making. Taken together, the influence priming had on reaction times but not on eye movements implies that the effect of priming is limited to categorization and decision making, while perceptual processes guide eye movements.

Conclusion

A new set of images with emerging properties has been created, and recognition performance and eye movements were measured. Our present study supports a perceptual account of target localization and recognition of EIs. Irrespective of priming, recognition was preceded by specific eye movement behavior with more fixations around the edges of the object. Moreover, observers who eventually recognized the object were inspecting its location already within the first second of viewing the image. Different types of priming did affect reaction time. This suggests that priming affected decision-making but not how visual stimuli are processed.

Having a more extensive and validated set of emergent stimuli provides opportunities for future studies. Separating the human ability to quickly detect and eventually recognize the complex emergent images in a robust way improves our understanding of human object recognition, perceptual and cognitive processes, and may aid the development of better biologically plausible computer vision algorithms.

Keywords: eye movements, viewing behavior, object recognition, emergence, gestalt, priming, saliency

Acknowledgments

BN was supported by a grant from the Netherlands Organisation for Scientific Research (NWO Brain & Cognition grant 433-09-233) to FWC. NG was supported by a scholarship from the (Chilean) National Commission for Scientific and Technological Research (BECAS CHILE & millennium center for neuroscience CENEM NC10 001 F). CP was supported by a scholarship from the Graduate School of Medical Sciences (GSMS) of the University Medical Center Groningen (UMCG). The authors would like to thank all partners within this project for their useful comments. In particular we thank Dr. Niloy J. Mitra, and Dr. Hung-Kuo Chu for sharing the program that was used to generate the emerging images and their help with the stimuli creation.

Commercial relationships: none
Corresponding author: Barbara Nordhjem.
Email: b.j.t.nordhjem@umcg.nl.
Address: University Medical Center Groningen, University of Groningen, Groningen, The Netherlands.

References

Biederman, I. (1972). Perceiving real-world scenes. *Science, 177*(4043), 77–80.

Biederman, I., & Cooper, E. E. (1991). Priming contour-deleted images: Evidence for intermediate representations in visual object recognition. *Cognitive Psychology, 23*(3), 393–419.

Bindemann, M. (2010). Scene and screen center bias early eye movements in scene viewing. *Vision*

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018

Research, 50(23), 2577–2587. doi:10.1016/j.visres.2010.08.016

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10,* 433–436.

Buswell, G. T. (1935). *How people look at pictures.* Chicago: University of Chicago Press.

Caldara, R., & Miellet, S. (2011). iMap: A novel method for statistical fixation mapping of eye movement data. *Behavior Research Methods, 43*(3), 864–878. doi:10.3758/s13428-011-0092-x

Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers, 34*(4), 613–617. doi:10.3758/BF03195489

Epshtein, B., & Ullman, S. (2005). Feature hierarchies for object classification. *International Conference on Computer Vision (Proc. IEEE), 1,* 220–227. doi:10.1109/ICCV.2005.98

Fiser, J., & Biederman, I. (2001). Invariance of long-term visual priming to scale, reflection, translation, and hemisphere. *Vision Research, 41*(2), 221–234.

Geisler, W., & Cormack, L. (2011). Models of overt attention. In S. P. Liversedge, I. D. Gilchrist, & S. Everling (Eds.), *Oxford handbook of eye movements* (pp. 439–454). New York: Oxford University Press.

Graf, P., & Schacter, D. L. (1985). Implicit and explicit memory for new associations in normal and amnesic subjects. *Journal of Experimental Psychology: Learning, Memory and Cognition, 11,* 501–518.

Harel, J., Koch, C., & Perona, P. (2007). Graph-based visual saliency. *Advances in Neural Information Processing Systems, 19,* 545–552. doi:10.1.1.70.2254

Ishikawa, T., & Mogi, K. (2011). Visual one-shot learning as an "anti-camouflage device": A novel morphing paradigm. *Cognitive Neurodynamics, 5*(3), 231–239. doi:10.1007/s11571-011-9171-z

Itti, L., Koch, C., & Niebur, E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 20*(11), 1254–1259.

Koch, C., & Ullman, S. (1985). Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, 4*(4), 219–227.

Kubilius, J., Wagemans, J., & Op de Beeck, H. P. (2011). Emergence of perceptual Gestalts in the human visual cortex: The case of the configural-superiority effect. *Psychological Science, 22*(10), 1296–1303. doi:10.1177/0956797611417000

Lee, H., & Op De. Beeck, H. P. (2012). Bistable Gestalts reduce activity in the whole of V1, not just the retinotopically predicted parts. *Journal of Vision, 12*(11):12, 1–14, doi:10.1167/12.11.12. [PubMed] [Article]

Lee, T. S. (2003). Computations in the early visual cortex. *Journal of Physiology (Paris), 97*(2), 121–139.

Malcolm, G. L., & Henderson, J. M. (2009). The effects of target template specificity on visual search in real-world scenes: Evidence from eye movements. *Journal of Vision, 9*(11):8, 1–13, doi:10.1167/9.11.8. [PubMed] [Article]

Marsman, J. B. C., Renken, R., Haak, K. V., & Cornelissen, F. W. (2013). Linking cortical visual processing to viewing behavior using fMRI. *Frontiers in Systems Neuroscience, 7*(12), 109. doi:10.3389/fnsys.2013.00109

Marsolek, C. J. (1999). Dissociable Neural Subsystems Underlie Abstract and Specific Object Recognition. *Psychological Science, 10*(2), 111–118. doi:10.1111/1467-9280.00117

Mensen, A., & Khatami, R. (2013). Advanced EEG analysis using threshold-free cluster-enhancement and non-parametric statistics. *NeuroImage, 67,* 111–118. doi:10.1016/j.neuroimage.2012.10.027

Mitra, N., Chu, H., Lee, T., & Wolf, L. (2009). Emerging images. *ACM Transactions on Graphics (Proc. SIGGRAPH), 28*(5), 163: 1–8. doi:10.1145/1618452.1618509

Nister, D., & Stewenius, H. (2006). Scalable recognition with a vocabulary tree. *Computer Society Conference on Computer Vision and Pattern Recognition (Proc. IEEE), 2,* 2161–2168. doi:10.1109/CVPR.2006.264.

Over, E. a B., Hooge, I. T. C., Vlaskamp, B. N. S., & Erkelens, C. J. (2007). Coarse-to-fine eye movement strategy in visual search. *Vision Research, 47*(17), 2272–2280. doi:10.1016/j.visres.2007.05.002

Palmeri, T. J., & Gauthier, I. (2004). Visual object understanding. *Nature Reviews. Neuroscience, 5*(4), 291–303. doi:10.1038/nrn1364

Pannasch, S., & Velichkovsky, B. M. (2009). Distractor effect and saccade amplitudes: Further evidence on different modes of processing in free exploration of visual images. *Visual Cognition, 17*(6-7), 1109–1131. doi:10.1080/13506280902764422

Pelli, D. G., Majaj, N. J., Raizman, N., Christian, C. J., Kim, E., & Palomares, M. C. (2009). Grouping in object recognition: The role of a Gestalt law in letter identification. *Cognitive Neuropsychology, 26*(1), 36–49. doi:10.1080/13546800802550134

Pernet, C. R., Chauveau, N., Gaspar, C., & Rousselet, G. a. (2011). LIMO EEG: A toolbox for hierarchical LInear MOdeling of ElectroEncephaloGraphic data. *Computational Intelligence and Neuroscience, 2011,* 831409. doi:10.1155/2011/831409

Potter, M. C. (1975). Meaning in visual search. *Science, 187*(4180), 965–966.

Schacter, D. L. (1992). Priming and multiple memory systems: Perceptual mechanisms of implicit memo-

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018

ry. *Journal of Cognitive Neuroscience, 4*(3), 244–256. doi:10.1162/jocn.1992.4.3.244

Schendan, H. E., Ganis, G., & Kutas, M. (1998). Neurophysiological evidence for visual perceptual categorization of words and faces within 150 ms. *Psychophysiology, 35*(3), 240–251.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(3), 411–426. doi:10.1109/TPAMI.2007.56

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage, 44*(1), 83–98. doi:10.1016/j.neuroimage.2008.03.061

Tatler, B. W. (2007). The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions. *Journal of Vision, 7*(14):4, 1–17, doi:10.1167/7.14.4. [PubMed] [Article]

Thorpe, S., Fize, D., & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381*(6582), 520–522. doi:10.1038/381520a0

Tonder, G. J. Van, & Ejima, Y. (2000). Bottom - up clues in target finding: Why a Dalmatian may be mistaken for an elephant. *Perception, 29*(2), 149–157. doi:10.1068/p2928

Unema, P. J. a., Pannasch, S., Joos, M., & Velichkovsky, B. M. (2005). Time course of information processing during scene perception: The relationship between saccade amplitude and fixation duration. *Visual Cognition, 12*(3), 473–494. doi:10.1080/13506280444000409

Velichkovsky, B., Joos, M., Helmert, J. R., & Pannasch, S. (2005). Two visual systems and their eye movements: Evidence from static and dynamic scene perception. In B. G. Bara, L. Barsalow, & M. Bucciarelli, (Eds.), *Proceedings of the XXVII conference of the Cognitive Science Society* (pp. 2283–2288). Mahwah, NJ: Lawrence Erlbaum.

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. a, Singh, M., & von der Heydt, R. (2012). A century of Gestalt psychology in visual perception: I. Perceptual grouping and figure-ground organization. *Psychological Bulletin, 138*(6), 1172–1217. doi:10.1037/a0029333

Yarbus, A. L. (1967). *Eye movements and vision*. New York: Plenum Press.

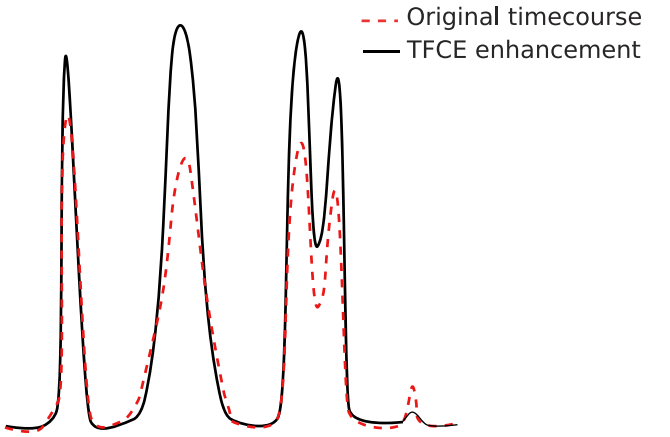

Figure S1. Illustration of the TFCE principle. A sharp peak is enhanced; a medium intensity peak with neighboring support is enhanced; two pairs are enhanced while a weak peak is suppressed.

Supplementary material

S1. The final 15 images that were used in this experiment were selected based on a pilot study with 30 images and 10 participants. Image selection from the pilot study was made on the basis of recognition time and accuracy. In order for an image to be included in this experiment, the average recognition time had to be above 3.5 s and recognized by at least 80% of the participants in the pilot study. By making hidden objects relatively large, we made sure recognition times were not prolonged due to their size.

S2. Eye movement traces were computed using a sample and hold technique. TFCE is used to transform the time courses followed by permutation statistics (1000 permutations). The TFCE principle is illustrated in in Figure S1 (adapted from Smith & Nichols, 2009).

The TFCE approach can be represented mathematically as $\mathrm{TFCE}(p) = \int_{h=h0}^{hp} e(h)^E h^H dh$

The TFCE value at a given point p is calculated as the integral of the cluster extend e multiplied with cluster heights h from $h0$ (the minimum value in the data) to h (the maximum value in the data). In the Matlab implementation, the TFCE scores were calculated as a sum using 1,000 steps. The two TFCE parameters, H and E, were here set to $H = 0$, $E = 1$, so that the local maximum in the TFCE time course will be represented proportionally to the local area under the curve. We refer to Smith and Nichols (2009) for an in-depth discussion of the TFCE parameters.

Downloaded From: http://jov.arvojournals.org/pdfaccess.ashx?url=/data/journals/jov/934201/ on 03/19/2018